# Using NOAA's New Climate Outlooks in Operational Hydrology

By Thomas E. Croley II[1]

**Abstract:** The National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center recently began issuing new multiple long-lead outlooks of meteorological probabilities. Operational hydrology approaches for generating probabilistic hydrological outlooks must be compatible with these meteorological outlooks yet preserve spatial and temporal relationships observed in past meteorology. Many approaches, however, either limit the use of historical data to be compatible with meteorological outlooks or limit compatibility with the outlooks to allow fuller use of historical data. An operational hydrology approach that uses all historical data while remaining compatible with many of the new long-lead outlooks, in order of user priority, is described here. The approach builds a hypothetical very large structured set of possible future scenarios, to be treated as a "sample" from which to estimate outlook probabilities and other parameters. The use of this hypothetical set corresponds to the weighted use of a scenario set based on historical data. The determination of weights becomes an optimization problem for the general case. An example illustrates the concepts and method.

## MAKING PROBABILISTIC OUTLOOKS

### Meteorological Probability Outlooks

Advances in long-range forecasting techniques recently enabled useable climate predictions beyond the previous 90-day limit. The National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center now provides each month a "Climate Outlook," consisting of a one-month outlook for the next month and 13 three-month outlooks, going into the future in overlapping fashion in one-month steps. Background and recent history on seasonal forecasting are provided elsewhere (Barnston et al. 1994; van den Dool 1994; Livezey 1990; Wagner 1989; Epstein 1988; Ropelewski and Halpert 1986; Gilman 1985).

The forecasts in the "Climate Outlook" are formed by a combination of methods. For U.S. air temperature and precipitation forecasts, these methods include: (1) Canonical correlation analysis (Barnston and Ropelewski 1992) relating spatial anomalies of sea surface temperature, Northern Hemisphere 700 mb height, and the U.S. surface climate; (2) use of observed interannual persistence of anomalies (Huang et al. 1994); and (3) forecasts from six-month general circulation models driven by sea surface temperatures [a set persisted from one half-month earlier and a set assembled from coupled ocean-atmosphere model runs (Ji et al. 1994)]. The general circulation model is a version of the National Meteorological Center medium range forecast model with special developmental emphasis on tropical processes.

Each outlook estimates probabilities of average air temperature and total precipitation falling within preselected value ranges. The value ranges (low, normal, and high) are defined as the lower, middle, and upper thirds of observations over the period 1961–90 for each variable. The climate outlooks presume that one of only four possibilities exist for the probabilities for each variable: (1) The probability of being in the high range exceeds one-third and the probability of being in the low range is reduced accordingly (it remains at one-third for the normal range), referred to as being "above normal"; (2) the probability of being in the normal range exceeds one-third and the probabilities of being in the low and high ranges are reduced accordingly and are equal, referred to as being "normal"; (3) the probability of being in the low range exceeds one-third and the probability of being in the high range is reduced accordingly (it remains at one-third for the normal range), referred to as being "below normal"; or (4) skill is insufficient to make a forecast and so probabilities of one-third in each range are used, referred to as "climatological."

### Hydrological Probability Outlooks

Users of these climate outlooks can interpret the forecast probabilities in terms of the impacts on themselves through "operational hydrology" approaches. Possibilities for the future are identified, which resemble past meteorology (preserving observed spatial and temporal relationships) yet are compatible with the climate outlooks. Some operational hydrology approaches consider historical meteorology as possibilities for the future by segmenting the historical record and using each segment with models to simulate a hydrological possibility for the future. Each segment of the historical record then has associated time series of meteorological and hydrological variables, representing a possible "scenario" for the future. The approach can then consider the resulting set of possible future scenarios as a statistical sample and infer probabilities and other parameters associated with both meteorology and hydrology through statistical estimation from this sample (Croley 1993; Croley and Lee 1993; Croley and Hartmann 1990; Day 1985; Smith et al. 1992). Other operational hydrology approaches use time series models of the historical data to generate the "sample." This increases the precision of the resulting statistical estimates, since large samples can be generated, but not the accuracy. Use of the historical record to directly build a sample for statistical estimation avoids the loss of representation consequent with the use of time series models, but requires a sufficiently large historical record.

The operational hydrology approach uses statistical sampling tools as if the set of possible future scenarios were a single "random sample" (i.e., the scenarios are independent of each other and equally likely). This means that the relative frequencies of selected events are fixed at values different (generally) from those specified in climate outlooks. Only by restructuring the set of possible future scenarios can we obtain relative frequencies of selected events that match climate outlooks. This restructuring violates the assumption of independent and equally likely scenarios (no random sample) from the point of view of the historical record (a priori information). However, the restructured set can be viewed as a random sample ("posterior" information) of scenarios conditioned on climate outlooks. There are many methods for restructuring the

[1]Res. Hydro., Great Lakes Envir. Res. Lab., 2205 Commonwealth Blvd., Ann Arbor, MI 48105-1593.

Note. Discussion open until December 1, 1996. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on August 24, 1995. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 1, No. 3, July, 1996. ©ASCE, ISSN 1084-0699/96/0003-0093–0102/$4.00 + $.50 per page. Paper No. 11449.

set of possible future scenarios (Croley 1993; Day 1985; Ingram et al. 1995; Smith et al. 1992).

## BUILDING A STRUCTURED SET

In building an operational hydrology set of possible future scenarios from which to estimate probabilities and other parameters associated with various meteorological and hydrological variables, consider constructing a structured set that, when treated as a statistical sample, guarantees that probability estimates for certain variables match a priori settings. That is, we can build a structured set of possible scenarios that gives relative frequencies of average air temperature and total precipitation (over various times in the scenarios) satisfying the a priori settings of the climate outlooks. We can arbitrarily construct a very large structured set of size $N$ by adding (duplicating) each of the available scenarios (in the original set of $n$ possible future scenarios); each scenario numbered $i$ ($i = 1, \ldots, n$) is duplicated $r_i$ times. By judiciously choosing these duplication numbers ($r_1, r_2, \ldots, r_n$), it is possible to force the relative frequency of any arbitrarily defined group of scenarios in the structured set to any desired value. For example, suppose only five of 50 (10%) 12-month scenarios beginning in June have an average June air temperature exceeding 30°C, and our a priori setting (from a climate outlook) for this exceedance is 20%. We could repeat each of these five scenarios nine times and repeat the other 45 scenarios four times to build a structured set. This structured set of size 225 (=5 × 9 + 45 × 4) would then have a relative frequency of 20% of average June air temperature exceeding 30°C (5 × 9)/225 = 0.2). For sufficiently large $N$, we can approximate a priori settings at any precision by using integer-valued duplication numbers, $r_i$. In addition

$$\sum_{i=1}^{n} r_i = N \tag{1}$$

The building of a structured set in this manner to match a priori settings is one of many arbitrary possibilities, but is suggested by considerations of constraints on estimated probability distribution for a single variable; see Appendix I.

By treating the $N$ scenarios in the very large structured set as a statistical sample, we can estimate probabilities and calculate other parameters for all variables. In particular, consider any variable $X$ (either historical meteorological or simulated hydrological); e.g., $X$ might be July-August-September total precipitation, end-of-August soil moisture storage, water surface temperature on day 55, or average June air temperature. We denote the event that a variable $X$ is less than or equal to a value $x$ as $\{X \le x\}$ and the probability of this event as $P[X \le x]$. This probability is estimated, when considering the very large structured set as a statistical sample, by the "relative frequency" of the event in the structured set. The relative frequency of event $\{X \le x\}$ is just the number of scenarios in which the event occurs divided by the set size $N$

$$\hat{P}[X \le x] = \sum_{k \in \Omega} \frac{1}{N}, \quad \Omega \equiv \{k | x_k^N \le x\} \tag{2}$$

where $\hat{P}[\,]$ = a probability estimate; and $x_k^N$ = value of variable $X$ for the $k$th scenario in the very large structured set of $N$ scenarios. [Read the set notation in (2) as "$\Omega$ is all values of $k$ such that $x_k^N \le x$."] Actually, there are only $n$ different values of $X$ ($x_i^n$, $i = 1, \ldots, n$) since these $n$ values were duplicated, each by a number $r_i$, to create $N$ values in the very large structured set. We can rewrite (2) in terms of the original set of possible future scenarios, for any variable $X$

$$\hat{P}[X \le x] = \sum_{i \in \Omega} \frac{r_i}{N}, \quad \Omega \equiv \{i | x_i^n \le x\} \tag{3}$$

Furthermore, we can write other estimators (defined over the large structured set of scenarios as if it was a statistical sample) in terms of the original set. Consider the $\gamma$-probability quantile for variable $X$, $\xi_\gamma$; it is defined by

$$P[X \le \xi_\gamma] = \gamma \tag{4}$$

The $\gamma$-probability quantile, $\xi_\gamma$, is estimated when considering the structured set as a statistical sample, by the $m$th order statistic, $y_m^N$, where $m = \gamma N$. Order all values of $X$ in the very large structured set ($x_k^N$, $k = 1, \ldots, N$) from smallest to largest to define the order statistics ($y_m^N$, $m = 1, \ldots, N$). The probability estimate is then

$$\hat{P}[X \le y_m^N] = \frac{m}{N}, \quad m = 1, \ldots, N \tag{5}$$

where $y_m^N = x_{k(m)}^N$; and $k(m)$ = number of the value in the structured set corresponding to the $m$th order. [For example, if the third value in the structured set, $x_3^N$, was the largest ($y_N^N = x_3^N$), then $k(N) = 3$]. Alternatively, (5) can be written as follows:

$$\hat{P}[X \le x_{k(m)}^N] = \sum_{i=1}^{m} \frac{1}{N}, \quad m = 1, \ldots, N \tag{6}$$

In terms of order statistics for the original set ($y_j^n$, $j = 1, \ldots, n$), there are $r_{i(j)}$ identical values of $y_j^n$ in the very large structured set where $i(j)$ is defined similarly to $k(m)$ but for the original set in which $j = 1, \ldots, n$ and $y_j^n = x_{i(j)}^n$. Eqs. (5) and (6) may be rewritten in terms of the original set of possible future scenarios (for any variable $X$)

$$\hat{P}[X \le y_j^n] = \hat{P}[X \le x_{i(j)}^n] = \sum_{l=1}^{j} \frac{r_{i(l)}}{N}, \quad j = 1, \ldots, n \tag{7}$$

Likewise, the sample mean and variance of variable $X$ over the structured set $\bar{x}$ and $S^2$, respectively, become, in terms of the original set

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k^N = \frac{1}{N} \sum_{i=1}^{n} r_i x_i^n \tag{8a}$$

$$S^2 = \frac{1}{N} \sum_{k=1}^{N} (x_k^N - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^{n} r_i (x_i^n - \bar{x})^2 \tag{8b}$$

Rewriting (3), (7), and (8)

$$\hat{P}[X \le x] = \frac{1}{n} \sum_{i \in \Omega} w_i, \quad \Omega \equiv \{i | x_i^n \le x\} \tag{9a}$$

$$\hat{P}[X \le y_j^n] = \frac{1}{n} \sum_{l=1}^{j} w_{i(l)}, \quad j = 1, \ldots, n \tag{9b}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} w_i x_i^n; \quad S^2 = \frac{1}{n} \sum_{i=1}^{n} w_i (x_i^n - \bar{x})^2 \tag{9c,d}$$

where

$$w_i = \frac{n}{N} r_i \tag{10}$$

Note that

$$\sum_{i=1}^{n} w_i = n \tag{11}$$

and if all $w_i = 1$, then (9) gives contemporary (unstructured) estimates from the original set, treated as a statistical sample. Other statistics can be similarly derived.

Eq. (9a) is functionally the same as that presented by Smith et al. (1992); here, the full development of statistic weights, including resampling and empirical distribution material, is

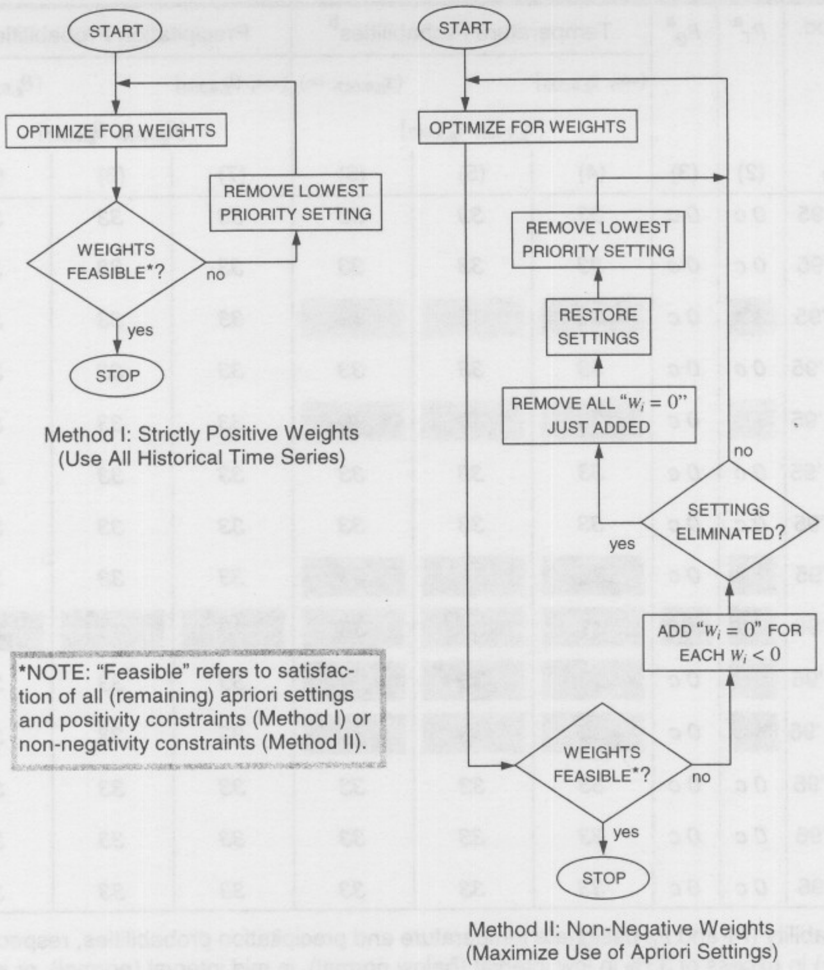


**FIG. 1. Procedural Algorithms for Determining Physically Relevant Weights**

**TABLE 1. Meteorological Quantiles on Lake Superior Basin[a] for Selected Periods**

| | Temperature Quantiles | | Precipitation Quantiles | |
|---|---|---|---|---|
| Period, $g$ | $\tau_{g,0.333}$ (°C) | $\tau_{g,0.667}$ (°C) | $\theta_{g,0.333}$ (mm) | $\theta_{g,0.667}$ (mm) |
| (1) | (2) | (3) | (4) | (5) |
| Jun | 13.38 | 14.43 | 69 | 106 |
| JJA | 15.18 | 16.29 | 242 | 295 |
| JAS | 14.49 | 15.12 | 240 | 299 |
| ASO | 10.32 | 11.18 | 253 | 282 |
| SON | 4.08 | 5.02 | 206 | 247 |
| OND | −3.40 | −2.09 | 178 | 216 |
| NDJ | −10.30 | −9.27 | 157 | 190 |
| DJF | −14.19 | −12.71 | 135 | 151 |
| JFM | −12.68 | −10.75 | 121 | 135 |
| FMA | −6.86 | −4.52 | 123 | 146 |
| MAM | 0.88 | 2.13 | 154 | 177 |
| AMJ | 8.03 | 8.55 | 197 | 230 |
| MJJ | 13.04 | 13.51 | 234 | 267 |

[a]Estimated from 1961–90 daily data over the Lake Superior Basin from 230 meteorological stations Thiessen averaged spatially (Croley and Hartmann 1985).

presented and extended for other statistics. Smith et al. used climatic indices from long-range forecasts to set their weights subjectively. Here, we will set the weights objectively to match a priori climate outlook probability settings. Appendix I contains an example for matching a single set of a priori settings of probabilities by finding appropriate values of the weights, $w_i$. A more general approach for matching multiple settings follows.

## CONSIDERING MULTIPLE OUTLOOKS

Now consider the case of multiple a priori settings (from a climate outlook) with which to match relative frequencies. For example, consider the settings from the new NOAA Climate Prediction Center ''Climate Outlook''

$$\hat{P}[T_g > \tau_{g,0.667}] = a_g, \quad g = 1, \ldots, 14 \tag{12a}$$

$$\hat{P}[T_g \le \tau_{g,0.333}] = b_g, \quad g = 1, \ldots, 14 \tag{12b}$$

$$\hat{P}[\tau_{g,0.333} < T_g \le \tau_{g,0.667}] = 1 - a_g - b_g, \quad g = 1, \ldots, 14 \tag{12c}$$

$$\hat{P}[Q_g > \theta_{g,0.667}] = c_g, \quad g = 1, \ldots, 14 \tag{12d}$$

$$\hat{P}[Q_g \le \theta_{g,0.333}] = d_g, \quad g = 1, \ldots, 14 \tag{12e}$$

$$\hat{P}[\theta_{g,0.333} < Q_g \le \theta_{g,0.667}] = 1 - c_g - d_g, \quad g = 1, \ldots, 14 \tag{12f}$$

where $T_g$ and $Q_g$ = average air temperature and total precipitation, respectively, over period $g$ ($g = 1$ corresponds to a one-month period, and $g = 2, \ldots, 14$ corresponds to 13 successive overlapping three-month periods); $\tau_{g,\gamma}$ and $\theta_{g,\gamma}$ = temperature and precipitation reference $\gamma$-probability quantiles for period $g$, respectively; and ($a_g$, $b_g$, $c_g$, and $d_g$, $g = 1, \ldots, 14$) = outlook settings. By definition, the reference $\gamma$-probability quantiles are estimated from the 1961–90 historical record for each period $g$. To illustrate (12), consider the June 1995 ''Climate Outlook''; there is a one-month June outlook ($g = 1$ or ''Jun'') and 13 three-month outlooks successively lagged by one month each ($g = 2$ or ''June-July-August'' or ''JJA,'' and $g$

| Period, $g$ | $P_T$[a] | $P_Q$[a] | Temperature Probabilities[b] | | | Precipitation Probabilities[b] | | |
|---|---|---|---|---|---|---|---|---|
| | | | $(-\infty, \tau_{g,0.333}]$ | $(\tau_{g,0.333}, \tau_{g,0.667}]$ | $(\tau_{g,0.667}, \infty)$ | $(-\infty, \theta_{g,0.333}]$ | $(\theta_{g,0.333}, \theta_{g,0.667}]$ | $(\theta_{g,0.667}, \infty)$ |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Jun '95 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |
| JJA '95 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |
| JAS '95 | 2 n | 0 c | 32 | 35 | 32 | 33 | 33 | 33 |
| ASO '95 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |
| SON '95 | 3 b | 0 c | 36 | 33 | 30 | 33 | 33 | 33 |
| OND '95 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |
| NDJ '95 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |
| DJF '95 | 1 a | 0 c | 32 | 33 | 34 | 33 | 33 | 33 |
| JFM '96 | 2 a | 10 b | 31 | 33 | 35 | 43 | 33 | 23 |
| FMA '96 | 1 a | 0 c | 32 | 33 | 34 | 33 | 33 | 33 |
| MAM '96 | 3 a | 0 c | 30 | 33 | 36 | 33 | 33 | 33 |
| AMJ '96 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |
| MJJ '96 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |
| JJA '96 | 0 c | 0 c | 33 | 33 | 33 | 33 | 33 | 33 |

[a]Probability ($P_T$ and $P_Q$ designate temperature and precipitation probabilities, respectively) in excess of 33% in low interval (below normal), in mid interval (normal), or in high interval (above normal); "no forecast" is indicated by "0 c" (climatological).

[b]Probabilities over the Climate Prediction Center's corresponding interval definitions. Probabilities expressed as percentages do not appear to sum to unity because of the two-digit round-off used here.

**FIG. 2. NOAA Climate Prediction Center June 1995 Climate Outlook Probabilities for Lake Superior Basin**

= 3, 4, . . . , 14 or ''JAS,'' ''ASO,'' . . . , ''JJA,'' respectively). Eqs. (12*c*) and (12*f*) are redundant with the rest of (12) because relative frequencies sum to unity

$$\hat{P}[T_g \le \tau_{g,0.333}] + \hat{P}[\tau_{g,0.333} < T_g \le \tau_{g,0.667}] + \hat{P}[T_g > \tau_{g,0.667}] = 1,\ g = 1, \ldots, 14 \tag{13a}$$

$$\hat{P}[Q_g \le \theta_{g,0.333}] + \hat{P}[\theta_{g,0.333} < Q_g \le \theta_{g,0.667}] + \hat{P}[Q_g > \theta_{g,0.667}] = 1,\ g = 1, \ldots, 14 \tag{13b}$$

Therefore, there are four independent settings in (12) for each of the 14 climate outlooks for a total of 56, if all outlooks are used.

Rewriting (12) and (13) in light of (9*a*)

$$\sum_{i \in A_g} w_i = a_g n,\ A_g \equiv \{i | t_{g,i} > \tau_{g,0.667}\},\ g = 1, \ldots, 14 \tag{14a}$$

$$\sum_{i \in B_g} w_i = b_g n,\ B_g \equiv \{i | t_{g,i} \le \tau_{g,0.333}\},\ g = 1, \ldots, 14 \tag{14b}$$

$$\sum_{i \in C_g} w_i = c_g n,\ C_g \equiv \{i | q_{g,i} > \theta_{g,0.667}\},\ g = 1, \ldots, 14 \tag{14c}$$

$$\sum_{i \in D_g} w_i = d_g n,\ D_g \equiv \{i | q_{g,i} \le \theta_{g,0.333}\},\ g = 1, \ldots, 14 \tag{14d}$$

$$\sum_{i=1}^{n} w_i = n \tag{14e}$$

where $t_{g,i}$ and $q_{g,i}$ = average air temperature and total precipitation, respectively, over period $g$ of scenario $i$. Alternatively, (14) can be written as follows:

$$\sum_{i=1}^{n} a_{k,i} w_i = e_k,\ k = 1, \ldots, 57 \tag{15}$$

where $a_{k,i}$ = 0 or 1 corresponding to the exclusion or inclusion, respectively, of each variable in the foregoing sets; and $e_k$ = climate outlook relative frequency settings specified in (12) times the number of available scenarios

$$e_k = a_k n,\quad k = 1, \ldots, 14 \tag{16a}$$

$$e_k = b_{k-14} n,\quad k = 15, \ldots, 28 \tag{16b}$$

$$e_k = c_{k-28} n,\quad k = 29, \ldots, 42 \tag{16c}$$

$$e_k = d_{k-42} n,\quad k = 43, \ldots, 56 \tag{16d}$$

$$e_k = n,\quad k = 57 \tag{16e}$$

Ordinarily, all of the Climate Prediction Center climate outlooks may not be used, in which case simply write (15) as

| Period, $g$[a] (1) | $k$[b] (2) | Interval[c] (3) | Inclusion in interval, $a_{k,i}$, $i = 1, \ldots, 45$[d] (4) | $e_k$[d] (5) |
|---|---|---|---|---|
| JAS '95 | 2 | $(\tau_{k,0.667}, \infty)$ | 110011010001110100100110010000000001000111010 | $0.32 \times 45$ |
| JAS '95 | 3 | $(-\infty, \tau_{k,0.333}]$ | 001100101000001001010001101001010010010000001 | $0.32 \times 45$ |
| SON '95 | 4 | $(\tau_{k,0.667}, \infty)$ | 100001101010111100001011010001000001100000000 | $0.30 \times 45$ |
| SON '95 | 5 | $(-\infty, \tau_{k,0.333}]$ | 000100000001000001100000101010011000011001010 | $0.36 \times 45$ |
| DJF '95 | 6 | $(\tau_{k,0.667}, \infty)$ | 100111110101100101001000001000011010001101111 | $0.34 \times 45$ |
| DJF '95 | 7 | $(-\infty, \tau_{k,0.333}]$ | 000000001010011010100001010011100100000000000 | $0.32 \times 45$ |
| JFM '96 | 8 | $(\tau_{k,0.667}, \infty)$ | 000111000100100100000000010001000101111001111 | $0.35 \times 45$ |
| JFM '96 | 9 | $(-\infty, \tau_{k,0.333}]$ | 010000000010001010000101010001100100000000000 | $0.31 \times 45$ |
| JFM '96 | 10 | $(\theta_{k,0.667}, \infty)$ | 111011100000000011100011001110100000000110000 | $0.23 \times 45$ |
| JFM '96 | 11 | $(-\infty, \theta_{k,0.333}]$ | 000000111111010000101000000010000110110001111 | $0.43 \times 45$ |
| FMA '96 | 12 | $(\tau_{k,0.667}, \infty)$ | 000101000100100000000000100010001011111001111 | $0.34 \times 45$ |
| FMA '96 | 13 | $(-\infty, \tau_{k,0.333}]$ | 010000000000000101001010100011000000000010000 | $0.32 \times 45$ |
| MAM '96 | 14 | $(\tau_{k,0.667}, \infty)$ | 001010100100010000010000100010001000111101111 | $0.36 \times 45$ |
| MAM '96 | 15 | $(-\infty, \tau_{k,0.333}]$ | 010001010001000010100111010000100000000010000 | $0.30 \times 45$ |
| Entire | 1 | $(-\infty, \infty)$ | 111111111111111111111111111111111111111111111 | $1.00 \times 45$ |

[a]Period as selected (highlighted) in Figure 2.

[b]Period renumbered by priority (1 ≡ highest) as in (17).

[c]Interval as defined in Table 1.

[d]Coefficients in (17) defined for each selected period, $k$, of the climate outlook, and for each scenario, $i$, in the historical record.

**FIG. 3. Boundary Condition Eq. (17) for June 1995 Outlook on Lake Superior**

$$\sum_{i=1}^{n} a_{k,i} w_i = e_k, \quad k = 1, \ldots, m \quad (17)$$

where $m \le 57$, and the appropriate equations, corresponding to the unused outlooks, are omitted. We must solve the equations in (17) simultaneously to find the weights.

Generally, $m \ne n$ and some of the equations may be either redundant or nonintersecting with the rest and must be eliminated. (If $m > n$, then $m - n$ of the equations must be either redundant or nonintersecting. This corresponds to not being able to simultaneously satisfy all climate outlook information with fewer scenarios than there are outlook boundary conditions.) Selection of some for elimination is facilitated by assigning each equation in (17) a priority reflecting its importance to the user. [The highest priority is given to the equation in (17) corresponding to (14*e*), guaranteeing that all relative frequencies sum to unity.] Each equation, in priority order starting with the next to highest priority, is compared to the set of all higher-priority equations and eliminated if it is redundant or does not intersect the set. By starting with the higher priorities, we ensure that each equation is compared with a known valid set of equations, and that we keep higher-priority equations in preference to lower-priority equations. Thus we can always reduce (17) so that $m \le n$. If $m = n$, then (17) can be solved via Gauss-Jordan elimination as a system of linear equations for the weights, $w_i$, since the equations are now independent and intersecting (in $n$-space). Otherwise, $m < n$, and (17) consists of the remaining independent intersecting equations.

There are multiple solutions to (17) for $m < n$, and the identification of the ''best'' set of weights requires the specification of a measure for comparing the solutions. One such measure is the deviation of the weights from unity, $\Sigma_{i=1}^{n} (w_i - 1)^2$. Solutions of (17) that give smaller values of this measure can be judged better than those that do not (and the resulting very large structured set of scenarios is more similar to the original set of scenarios in this sense). Other measures are also possible, including those using other functions expressing deviation of the weights from a goal, or measures defined on the resulting joint probability distribution function estimates (looking at similarity in joint distributions between the very large structured set and the original set). Here, it is judged desirable to be as similar to the original set as possible, in terms of relative frequencies of the selected events.

We can formulate an optimization problem to minimize the foregoing deviation of weights from unity in selecting a solution to (17)

**TABLE 2. Climate Outlook Weights Using All Historical Time Series**[a]

| Year (1) | Weight (2) | Year (3) | Weight (4) | Year (5) | Weight (6) |
|---|---|---|---|---|---|
| 1948 | 0.444378 | 1963 | 0.259718 | 1978 | 1.527387 |
| 1949 | 1.659873 | 1964 | 1.527387 | 1979 | 1.112034 |
| 1950 | 1.089694 | 1965 | 1.112034 | 1980 | 1.459070 |
| 1951 | 0.927374 | 1966 | 1.183255 | 1981 | 1.527387 |
| 1952 | 0.150880 | 1967 | 1.089694 | 1982 | 0.157130 |
| 1953 | 0.259718 | 1968 | 0.982324 | 1983 | 1.007623 |
| 1954 | 0.450628 | 1969 | 1.659873 | 1984 | 1.545569 |
| 1955 | 0.335539 | 1970 | 1.192282 | 1985 | 1.675279 |
| 1956 | 0.528100 | 1971 | 1.104530 | 1986 | 1.459070 |
| 1957 | 0.688826 | 1972 | 1.675279 | 1987 | 0.335539 |
| 1958 | 1.636225 | 1973 | 1.098279 | 1988 | 1.083444 |
| 1959 | 1.105783 | 1974 | 1.112034 | 1989 | 0.921124 |
| 1960 | 0.259718 | 1975 | 1.621390 | 1990 | 0.688826 |
| 1961 | 0.521850 | 1976 | 1.536542 | 1991 | 0.921124 |
| 1962 | 1.104530 | 1977 | 1.104530 | 1992 | 0.157130 |

[a]Solution of Eq. (18) with Fig. 3 coefficients and Method 1 in Fig. 1; a priori settings for JAS, SON, DJF, and JFM temperature probabilities are used and settings for FMA and MAM temperature probabilities and JFM precipitation probabilities are unused.

**TABLE 3. Climate Outlook Weights Using All A Priori Climate Settings**[a]

| Year (1) | Weight (2) | Year (3) | Weight (4) | Year (5) | Weight (6) |
|---|---|---|---|---|---|
| 1948 | 0 | 1963 | 0.450000 | 1978 | 1.269962 |
| 1949 | 1.060486 | 1964 | 1.269962 | 1979 | 1.919873 |
| 1950 | 0.312190 | 1965 | 0.424136 | 1980 | 1.813411 |
| 1951 | 1.008031 | 1966 | 1.808557 | 1981 | 1.279712 |
| 1952 | 0 | 1967 | 1.879379 | 1982 | 0.171944 |
| 1953 | 0 | 1968 | 1.912046 | 1983 | 0.911242 |
| 1954 | 0 | 1969 | 2.627675 | 1984 | 1.795797 |
| 1955 | 0.357372 | 1970 | 0 | 1985 | 1.875076 |
| 1956 | 1.137376 | 1971 | 0.379306 | 1986 | 1.884862 |
| 1957 | 0.977323 | 1972 | 1.803624 | 1987 | 0 |
| 1958 | 1.355692 | 1973 | 1.724416 | 1988 | 1.737354 |
| 1959 | 1.264911 | 1974 | 0.424136 | 1989 | 0.767599 |
| 1960 | 0.025845 | 1975 | 1.297178 | 1990 | 0.977323 |
| 1961 | 0.825493 | 1976 | 0.366735 | 1991 | 0.839051 |
| 1962 | 0.460508 | 1977 | 2.522282 | 1992 | 0.082140 |

[a]Solution of Eq. (18) with Fig. 3 coefficients and method 2 in Fig. 1; all a priori settings in Fig. 3 are used.

$$\min \sum_{i=1}^{n} (w_i - 1)^2; \quad \text{subject to } e_k \sum_{i=1}^{n} a_{k,i} w_i = e, \; k = 1, \ldots, m \tag{18}$$

By defining the Lagrangian for this problem (Hillier and Lieberman 1969)

$$L = \sum_{i=1}^{n} (w_i - 1)^2 - \sum_{k=1}^{m} \lambda_k \left( \sum_{i=1}^{n} a_{k,i} w_i - e_k \right) \tag{19}$$

(where $\lambda_k$ = unit penalty of violating the $k$th constraint in the optimization) and by setting the first derivatives of the Lagrangian with respect to each variable to zero

$$\frac{\partial L}{\partial w_i} = 2(w_i - 1) - \sum_{k=1}^{m} \lambda_k a_{k,i} = 0, \quad i = 1, \ldots, n \tag{20a}$$

$$\frac{\partial L}{\partial \lambda_k} = -\sum_{i=1}^{n} a_{k,i} w_i + e_k = 0, \quad k = 1, \ldots, m \tag{20b}$$

we have a set of necessary but not sufficient conditions for the problem of (18). Eqs. (20*a*,*b*) are linear and solvable via the Gauss-Jordan method of elimination. Sufficiency may be checked by inspection of the solution space in the vicinity of the solution.

The solution of (18) may give positive, zero, or negative weights, but only nonnegative weights make physical sense and we must further constrain the optimization to nonnegative weights. This can be done by introducing nonnegativity inequality constraints into (18), converting them to equality constraints by defining additional variables, redefining the Lagrangian in (19) in terms of these additional constraints and variables, and determining the corresponding additional equations in (20). These additional equations would require enumeration of all zero points or "roots" of (20) (a root is a solution with zero-valued weights). However, this is computationally impractical since it can involve the inspection of many roots [e.g., for $n = 50$, there are $2^{50} - 1$ roots ($>10^{15}$)]. Furthermore, nonnegativity constraints can result in infeasibility (there is no solution). In this case, additional lowest priority equations must be eliminated from (17) to allow a nonnegative solution. The smallest number possible should be eliminated so that as many of the a priori settings as possible are preserved. Elimination of equations can proceed in a variety of manners. If higher-priority equations were eliminated, it might be possible to eliminate fewer equations. This would involve further assessment of the importance of a small set of high-priority equations versus a larger set of lower priority equations, which is impossible to make in a general manner for all situations. The following two methods provide systematic procedures for finding nonnegative weights through the elimination of lowest-priority equations. They also avoid the direct use of nonnegativity constraints in (18), thus avoiding inspection of the large number of roots that can result.

The first method guarantees that only strictly positive weights will result; this means that all possible future scenarios are used (no scenario is weighted by zero and effectively eliminated) in estimating probabilities and other parameters. The procedure is to solve (18) without additional "positivity" constraints (all weights are positive). If the solution also satisfies the positivity constraints, then we also have a solution to the further-constrained optimization problem, and we are finished. If the solution does not satisfy all the positivity constraints, then it cannot be an actual solution to the further-constrained problem. This indicates some positivity constraints are active in the actual solution and the constrained optimum may exist only in the limit as some of the weights approach zero (nonpositive). We need not solve this further-constrained problem, since that solution does not interest us. Instead, we remove the lowest-priority equation (reduce $m$ by one) in (17) and (18) and resolve the optimization, repeating until we have a strictly positive solution. Fig. 1 summarizes the procedural algorithm for this method.

Alternatively, if we are willing to disallow some of the possible future scenarios (allow zero weights), then we can strive to satisfy more of the a priori settings [more of the equations in (17)] in the solution. In the second method, if negative weights are observed in the solution of (18), we add zero constraints ($w_i = 0$), corresponding only to those weights that are negative, and solve this further-constrained problem. However, introducing selected zero constraints will either eliminate some a priori settings [equations in (17)] (because of infeasibility but not because of redundancy) or it will not. If it does, the solution to the further-constrained problem cannot be feasible in the predecessor problem. The method instead removes the lowest-priority constraint in the predecessor problem (reduce $m$ by one) and resolves the optimization. It it does not (eliminate some a priori settings), then the optimum solution to the further-constrained problem is feasible (and optimum) in the predecessor problem, but new negative weights could be generated. If no negative weights are generated then we are finished. If some negative weights are generated, the process (of adding selected zero constraints and solving the further-con-

**TABLE 4. June 1995 Lake Superior Outlook of Monthly Total Net Basin Supply (mm)[a]**

| Month (1) | Quantiles 1% (2) | 5% (3) | 10% (4) | 20% (5) | 50% (6) | 80% (7) | 90% (8) | 95% (9) | 99% (10) | Mean (11) | Standard deviation (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| June 1995 | 88 | 99 | 103 | 108 | 149 | 167 | 185 | 188 | 198 | 141 | 30 |
| July 1995 | 68 | 80 | 92 | 101 | 114 | 142 | 153 | 166 | 180 | 120 | 26 |
| August 1995 | 22 | 44 | 55 | 82 | 95 | 131 | 137 | 151 | 183 | 102 | 35 |
| September 1995 | −14 | −5 | 1 | 39 | 65 | 109 | 137 | 157 | 176 | 75 | 47 |
| October 1995 | −14 | −5 | 7 | 23 | 46 | 77 | 89 | 93 | 102 | 49 | 30 |
| November 1995 | −58 | −42 | −18 | −14 | 2 | 30 | 59 | 66 | 86 | 10 | 33 |
| December 1995 | −65 | −59 | −50 | −39 | −28 | −15 | −1 | 2 | 16 | −26 | 18 |
| January 1996 | −77 | −65 | −50 | −40 | −23 | −15 | 6 | 8 | 13 | −25 | 20 |
| February 1996 | −55 | −37 | −27 | −22 | −14 | 13 | 21 | 26 | 58 | −6 | 23 |
| March 1996 | −27 | −25 | −7 | 5 | 21 | 59 | 82 | 92 | 115 | 34 | 36 |
| April 1996 | 41 | 62 | 75 | 87 | 120 | 151 | 164 | 173 | 177 | 121 | 32 |
| May 1996 | 94 | 100 | 104 | 127 | 159 | 192 | 228 | 234 | 246 | 162 | 42 |

[a]Forecast nonexceedance quantiles, mean, and standard deviation are expressed as overlake depths. The quantiles are interpolated from Eq. (9*b*) and the mean and standard deviation are from Eq. (9*c,d*), with the weights from Table 3. This hydrological outlook corresponds to the Climate Prediction Center "Climate Outlook" for June 1995, using probability settings on temperature for periods JAS, SON, DJF, JFM, FMA, and MAM, and on precipitation for the JFM period.

strained problem) can be repeated either until an optimum solution is generated to the further-constrained problem that is nonnegative or until a priori settings are eliminated. If the latter occurs, the method removes the lowest-priority constraint in the predecessor problem (reduce $m$ by one) and resolves the optimization. This process is repeated until we have a nonnegative solution. Fig. 1 also summarizes the procedural algorithm for this method.

## EXAMPLE CONSIDERATION OF MULTIPLE OUTLOOKS

Consider the following example. The Great Lakes Environmental Research Laboratory (GLERL) hydrology models are to be used to estimate the 12-month probabilistic outlook of net basin supply for Lake Superior beginning June 1995 by using the NOAA Climate Prediction Center "Climate Outlook" for June 1995. (Net basin supply is the algebraic sum of overlake precipitation, lake evaporation, and basin runoff to the lake.) The outlook will be made by identifying all 12-month meteorological time series that start in June from the available historical record of 1948–93; there are 45 such times series for each meteorological variable. The time series for all meteorological variables will be used in simulations with GLERL's hydrology models and current initial conditions to estimate the 45 associated time series for each hydrological variable. Each set of historical meteorological and associated hydrological time series, corresponding to each segment of the historical record, represent a possible future scenario. The 45 scenarios will be used as a statistical sample in an operational hydrology approach to make the probabilistic outlook. We will incorporate the Climate Prediction Center "Climate Outlook" by using selected period outlook settings as boundary conditions in the determination of weights to apply to our scenario set. We use these weights, through estimates from (9), to make our probabilistic outlook.

We must begin by abstracting historical quantiles of air temperature and precipitation for the Lake Superior basin; these are presented in Table 1 for the periods of interest in making the June outlook. These were estimated from the 1961–90 period in accordance with definitions provided by the Climate Prediction Center for use of their climate outlooks. These quantile estimates are the basis for interpretation of the Climate Prediction Center's climate outlooks.

The NOAA Climate Prediction Center "Climate Outlook" for June 1995 (made May 18, 1995) over the Lake Superior Basin is given in Fig. 2 in columns two and three. They are interpreted, in accordance with specifications of the Climate Prediction Center [and as described in the section on "Meteorological Probability Outlooks" and in the previous section; see (12)], to construct the probabilities associated with the reference quantiles in Table 1; these are given in columns four through nine in Fig. 2. The shaded entries in Fig. 2 denote outlook probabilities designated as significant by the Climate Prediction Center, who suggest that the remainder be estimated from climatology since they have insufficient skill to make outlooks in those cases.

The highlighted entries in Fig. 2 are used arbitrarily, in priority of their appearance, to make the hydrological outlook. These seven outlook settings and the reference quantiles in Table 1 are used with inspection of all 45 scenarios to construct the 15 equations represented by (17) in Fig. 3. Table 2 presents the solution of these equations, found by minimizing the deviation of weights from unity, as in (18), by using the first procedural algorithm in Fig. 1 (using all scenarios). While all 45 scenarios are used (all weights are strictly positive), not all of the selected a priori climate settings can be used. The temperature probability settings for JAS, SON, DJF, and JFM were used while the temperature probability settings for FMA and MAM and the precipitation probability setting for JFM were unused.

Table 3 presents the solution of the equations with coefficients in Fig. 3, found by minimizing the deviation of weights with unity, as in (18), by using the second procedural algorithm in Fig. 1 (maximizing use of the a priori climate outlook settings). All seven a priori climate settings, highlighted in Fig. 2, can be included. Table 3 shows that six weights were assigned values of zero to enable this inclusion. This means that the scenarios starting in June 1948, 1952, 1953, 1954, 1970, and 1987 are unused in the ensuing probabilistic outlook.

Finally, as an example for one hydrological variable, the probabilistic outlook for net basin supply (NBS), over the 12 months from June 1995 through May 1996, is given in Table 4. There were 45 values of monthly NBS, corresponding to the 45 scenarios used in the simulation, for each of the 12 months. Each value was multiplied by its respective weight from Table 3, as in (9), to compute various statistics for the probabilistic outlook each month. Selected quantiles from the forecast NBS probability distribution and the mean and standard deviation for each month of the outlook are displayed in Table 4. Since the weights of Table 3 were used, the probabilistic outlook in Table 4 represents use of all selected a priori climate outlook settings.

## CONCLUSIONS

The operational hydrology approach described here uses all (method 1) or most (method 2) historical information while preserving many of the long-term meteorological probability outlooks provided by NOAA's Climate Prediction Center. Some other approaches severely limit the use of historical data to be compatible with climate outlooks or use all historical data only by ignoring these outlooks. The use of a hypothetical very large structured set of scenarios (matching climate outlooks) to estimate hydrological outlook probabilities corresponds to the use of the weighted original set of possible future scenarios estimated from the historical record. (Each scenario consists of an actual segment of the historical meteorological record and its associated hydrological transformation made with appropriate models.) The building of this hypothetical very large structured set is an arbitrary concept that was useful in defining the weights. The National Weather Service is now considering weighting methods for their Extended Streamflow Prediction (ESP) operational hydrology approach (Day 1985; Smith et al. 1992) that couple historical time series of precipitation with precipitation forecasts (Ingram et al. 1995).

Still other approaches use time series models, fit to historical data, to generate a large sample, increasing precision but not accuracy in the resulting statistical estimates. Direct use of the historical record to build a sample avoids the loss of representation consequent with time series models. In addition, it may not be clear how to modify time series models to agree with climatic outlooks and still be representative of the underlying behavior originally captured in the time series models. Nevertheless, if time series models are used in building the sample, weighting of this sample, in the manner described here, to agree with climatic outlooks is straightforward and still could be used.

The determination of these weights involves several choices also made arbitrarily here. For example, the weights could be determined directly from multiple climate outlooks, as exemplified in Appendix I for a single climate outlook. This would involve restrictions on the multiple climate outlooks not considered in this paper. The formulation of an optimization problem, used here, allows for a more general approach in determining these weights in the face of multiple outlooks. However, this formulation also involves arbitrary choices, the largest of which is the selection of a relevant objective function. As mentioned earlier, other measures of relevance of the weights to a goal are possible and could require reformulation of the solution methodology. An early approach, not reported here, minimized the sum of squared differences between the relative frequencies associated with the bivariate distribution of precipitation and temperature before and after application of the weights. The goal was to make the resulting joint distribution as similar as possible to that observed historically while making the marginal distributions match the climate outlooks. Unfortunately, the method was intractable for consideration of more than one climate outlook.

Also not reported in this paper was an effort where consideration was made of linear objective functions; the weights were linearly related to a goal of making them as close as possible to unity. This was an effort to make the optimization problem amenable to linear programming solution methodologies. That way, additional constraints on the weights for positivity or nonnegativity could be added directly to the optimization and evaluated systematically. The Simplex method (Wagner 1975) was used to solve the resulting linear optimization problem. However, the large number of roots consequent in practical problems for a nonunique optimum still rendered the solution computationally intractable. Nevertheless, this formulation could be used in the manner described for the solution to (20) (where positivity or nonnegativity constraints are considered systematically outside of the optimizations) without loss of generality, if a linear objective function was deemed more suitable in an application.

An important advantage associated with the computation of a weighted sample in the operational hydrology approach described here, and as with ESP, is the independence of the weights and the hydrology models. After model simulations are made to build a set of possible future scenarios for analysis, several probabilistic outlooks can be generated with weights corresponding to the use of different climate outlooks, different methods of considering the climate outlooks, and alternative selections of just which of the 14 outlooks that are available each month to use. In making these alternative analyses and weights (re)computations, it is unnecessary to redo the model simulations to rebuild the set. This is a real savings when the model simulations are extensive, as is the case with Great Lakes hydrological outlooks. This also enables efficient consideration of other ways of using the weights to make probabilistic outlooks. For example, our use of nonparametric statistics in (9) restricts the range of any variable to that present in the historical record or in their hydrological transformations. An alternative that does not restrict range in this manner is to hypothesize a distribution family (e.g., normal, log-normal, log-Pearson type III) and to estimate its moments by using sample statistics defined analogously to those in (9). The detractor for parametric estimation is hypothesizing the family of distributions to use.

Most significantly, the method allows joint consideration of multiple meteorological outlooks defined over different lengths and periods of time. It can be easily extended to incorporate consideration of six- to 14-day outlooks, for which there is relatively greater skill, as well as other period outlooks.

Computer code is available, to make all computations (outside of the hydrological modeling), for use by others in utilizing the NOAA Climate Prediction Center ''Climate Outlook.'' The code finds all necessary reference quantiles, for using a climate outlook, from a user-supplied file of historical daily air temperature and precipitation, sets up all equations in (17), formulates the optimization problem of (18), and performs the sequential optimizations [solutions of (20)] with either method in Fig. 1 (either use all historical data or maximize use of a priori climate outlook settings). The code is available both as a stand-alone FORTRAN implementation, for use under a variety of operating systems, and as a specially designed user interface Windows application. The latter also allows readily understandable user interpretation of the NOAA Climate Prediction Center's ''Climate Outlooks'' and easy user assignment of relevant priorities.

## ACKNOWLEDGMENTS

Discussions with Lynn Herche focused the resampling and empirical distribution material. This is GLERL contribution number 969.

## APPENDIX I. ALTERNATIVE CONSIDERATION OF A SINGLE CLIMATE OUTLOOK

Consider probability estimates for a single variable that match a priori settings. For example, suppose that our a priori settings for average temperature during the June-July-August climate outlook (or JJA) $T_{JJA}$ are a 38.3% chance of exceeding the 66.7% quantile (determined for JJA within 1961–90) $\tau_{JJA,0.667}$, a 28.3% chance of not exceeding the 33.3% quantile $\tau_{JJA,0.333}$, and a 33.4% chance of being between the two

$$\hat{P}[T_{JJA} > \tau_{JJA,0.667}] = 0.383, \text{ over the upcoming outlook period} \tag{21a}$$

$$\hat{P}[T_{JJA} \leq \tau_{JJA,0.333}] = 0.283, \text{ over the upcoming outlook period} \tag{21b}$$

$\hat{P}[\tau_{JJA,0.333} < T_{JJA} \leq \tau_{JJA,0.667}] = 0.334,$

over the upcoming outlook period (21c)

where $\hat{P}$ = relative frequency, used as a probability estimate; and the quantiles are defined from historical data

$$\hat{P}[T_{JJA} \leq \tau_{JJA,0.667}] = 0.667, \text{ over the historical 1961–90 period} \quad (22a)$$

$$\hat{P}[T_{JJA} \leq \tau_{JJA,0.333}] = 0.333, \text{ over the historical 1961–90 period} \quad (22b)$$

We will construct a very large structured set, of size $N$, of scenarios with relative frequencies satisfying (21) by duplicating original scenarios, such that

$$\frac{N_U}{N} = \hat{P}[T_{JJA} > \tau_{JJA,0.667}] = 0.383 \quad (23a)$$

$$\frac{N_L}{N} = \hat{P}[T_{JJA} \leq \tau_{JJA,0.333}] = 0.283 \quad (23b)$$

$$\frac{N - N_U - N_L}{N} = \hat{P}[\tau_{JJA,0.333} < T_{JJA} \leq \tau_{JJA,0.667}] = 0.334 \quad (23c)$$

where $N_U$ = number of scenarios with $T_{JJA} > \tau_{JJA,0.667}$; and $N_L$ = number of scenarios with $T_{JJA} \leq \tau_{JJA,0.333}$. The original sample of $n$ scenarios has $n_U$ scenarios with $T_{JJA} > \tau_{JJA,0.667}$ and $n_L$ scenarios with $T_{JJA} \leq \tau_{JJA,0.333}$. Each of the $n_U$ scenarios will be duplicated $N_U/n_U$ times and each of the $n_L$ scenarios will be duplicated $N_L/n_L$ times. By making the structured set sufficiently large, the approximations in (23) can be made as close as desired. In the limit, as the integers $N$, $N_U$, and $N_L$ grow, the approximations in (23) approach equalities.

Of the original $n$ scenarios, the $i$th scenario is repeated $r_i$ times, where

$$r_i = \frac{N_L}{n_L}, \forall\, i \,|\, t_{JJA,i} \leq \tau_{JJA,0.333} \quad (24a)$$

$$r_i = \frac{N_U}{n_U}, \forall\, i \,|\, t_{JJA,i} > \tau_{JJA,0.667} \quad (24b)$$

$$r_i = \frac{N - N_U - N_L}{n - n_U - n_L}, \forall\, i \,|\, \tau_{JJA,0.333} < t_{JJA,i} \leq \tau_{JJA,0.667} \quad (24c)$$

where $t_{JJA,i}$ = average JJA air temperature in scenario $i$. For $N$ sufficiently large, each ratio, $r_i$, is an integer if the probability estimate settings are specified only to a fixed number of digits. Statistics can be written as functions of either the very large structured set $(x_1^N, \ldots, x_N^N)$, or the original set $(x_1^n, \ldots, x_n^n)$. For example, the structured sample mean and variance, $\bar{x}$ and $S^2$, respectively, are

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k^N = \frac{1}{N} \sum_{i=1}^{n} r_i x_i^n = \frac{1}{n} \sum_{i=1}^{n} w_i x_i^n \quad (25a)$$

$$S^2 = \frac{1}{N} \sum_{k=1}^{N} (x_k^N - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^{n} r_i (x_i^n - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n} w_i (x_i^n - \bar{x})^2 \quad (25b)$$

where

$$w_i = \frac{n}{N} r_i = 0.283 \frac{n}{n_L}, \forall\, i \,|\, t_{JJA,i} \leq \tau_{JJA,0.333} \quad (26a)$$

$$w_i = \frac{n}{N} r_i = 0.383 \frac{n}{n_U}, \forall\, i \,|\, t_{JJA,i} > \tau_{JJA,0.667} \quad (26b)$$

$$w_i = \frac{n}{N} r_i = 0.334 \frac{n}{n - n_U - n_L}, \forall\, i \,|\, \tau_{JJA,0.333} < t_{JJA,i} \leq \tau_{JJA,0.667} \quad (26c)$$

If the period 1961–90 was also our entire historical record then, by definition, $n_L/n = 0.333$ and $n_U/n = 0.333$. Therefore

$$w_i = 0.283/0.333 = 0.850, \forall\, i \,|\, t_{JJA,i} \leq \tau_{JJA,0.333} \quad (27a)$$

$$w_i = 0.383/0.333 = 1.150, \forall\, i \,|\, t_{JJA,i} > \tau_{JJA,0.667} \quad (27b)$$

$$w_i = 0.334/0.334 = 1.000, \forall\, i \,|\, \tau_{JJA,0.333} < t_{JJA,i} \leq \tau_{JJA,0.667} \quad (27c)$$

Other statistics can be similarly derived. Furthermore, the preceding development can be made for variables besides temperature and for any period other than JJA without loss of generality, including single-month periods. It is also possible to define alternative very large structured sets based on other probability quantiles besides the two used here, 33.3% and 66.7%, and on other systematic manners of duplicating the original scenarios.

## APPENDIX II. REFERENCES


Barnston, A. G., and Ropelewski, C. F. (1992). "Prediction of ENSO episodes using canonical correlation analysis." *J. of Climate*, 5(11), 1316–1345.

Barnston, A. G. et al. (1994). "Long-lead seasonal forecasts—where do we stand?" *Proc., Seasonal Forecasting for the Workshop on Long-Lead Climate Forecasts*, Nat. Oceanic and Atmospheric Admin. (NOAA), Washington, D.C.

Croley, T. E. II (1993). "Probabilistic Great Lakes hydrology outlooks." *Water Resour. Bull.*, 29(5), 741–753.

Croley, T. E. II, and Hartmann, H. C. (1985). "Resolving Thiessen Polygons." *J. Hydro.*, Amsterdam, The Netherlands, Vol. 76, 363–379.

Croley, T. E. II, and Hartmann, H. C. (1990). "GLERL's near real-time hydrological outlook package." *Proc., Great Lakes Water Level Forecasting and Statistics Symp.*, Great Lakes Commission, Ann Arbor, Mich., 63–72.

Croley, T. E. II, and Lee, D. H. (1993). "Evaluation of Great Lakes net basin supply forecasts." *Water Resour. Bull.*, 29(2), 267–282.

Day, G. N. (1985). "Extended streamflow forecasting using NWSRFS." *J. Water Resour. Plng. and Mgmt.*, ASCE, Vol. 111, 157–170.

Epstein, E. S. (1988). "Long-range weather prediction: limits of predictability and beyond." *Weather and Forecasting*, 3(1), 69–75.

Gilman, D. L. (1985). "Long-range forecasting: the present and the future." *Bull. Am. Meteorological Soc.*, 66(2), 159–164.

Hillier, F. S., and Lieberman, G. J. (1969). *Introduction to operations research. Appendix 2: classical optimization techniques.* Holden-Day, San Francisco, Calif., 603–608.

Huang, J., van den Dool, H. M. and Barnston, A. G. (1994). "Long-lead seasonal temperature prediction using optimal climate normals." *Proc., Seasonal Forecasting for Workshop on Long-Lead Forecasts*, Nat. Oceanic and Atmospheric Admin. (NOAA), Washington, D.C.

Ingram, J. J., Hudlow, M. D., and Fread, D. L. (1995). "Hydrometeorological coupling for extended streamflow predictions." *Proc., Conf. on Hydro.; 75th Annu. Meeting of Am. Meteorological Soc.*, Am. Meteorological Soc., Boston, Mass., 186–191.

Ji, M., Kumar, A., and Leetmaa, A. (1994). "A multiseason climate forecast system at the National Meteorological Center." *Bull. Am. Meteorological Soc.*, 75(4), 569–577.

Livezey, R. E. (1990). "Variability of skill of long-range forecasts and implications for their use and value." *Bull. Am. Meteorological Soc.*, 71(3), 300–309.

Ropelewski, C. F., and Halpert, M. S. (1986). "North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO)." *Monthly Weather Rev.*, 114(12), 2352–2362.

Smith, J. A., Day, G. N., and Kane, M. D. (1992). "Nonparametric framework for long-range streamflow forecasting." *J. Water Resour. Plng. and Mgmt.*, ASCE, 118(1), 82–92.

van den Dool, H. M. (1994). "Long-range weather forecasts through numerical and empirical methods." *Dynamics of Atmospheres and Oceans*, Vol. 20, 247–270.

Wagner, H. M. (1975). "Simplex method of solution." *Principles of operations research*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 91–167.

Wagner, A. J. (1989). "Medium- and long-range forecasting." *Weather and Forecasting*, 4(3), 413–426.

## APPENDIX III. NOTATION

*The following symbols are used in this paper:*

$A_g$ = set of indices of scenarios containing average air temperature for period $g$ in the upper third of its 1961–90 range;
$a_g$ = a priori climate outlook probability setting for average air temperature for period $g$ in the upper third of its 1961–90 range;
$a_{k,i}$ = coefficient in $k$th equation on $i$th weight (for $i$th scenario) in Eqs. (15), (17), (18), (19), and (20);
$B_g$ = set of indices of scenarios containing average air temperature for period $g$ in the lower third of its 1961–90 range;
$b_g$ = a priori climate outlook probability setting for average air temperature for period $g$ in the lower third of its 1961–90 range;
$C_g$ = set of indices of scenarios containing average precipitation for period $g$ in the upper third of its 1961–90 range;
$c_g$ = a priori climate outlook probability setting for average precipitation for period $g$ in the upper third of its 1961–90 range;
$D_g$ = set of indices of scenarios containing average precipitation for period $g$ in the lower third of its 1961–90 range;
$d_g$ = a priori climate outlook probability setting for average precipitation for period $g$ in the lower third of its 1961–90 range;
$e_k$ = selected weights sum limit in $k$th Eq. in (15), (17), (18), (19), and (20), corresponding to an a priori climate outlook probability setting;
$L$ = objective function (the Lagrangian) for an unconstrained optimization reformulated from the objective function for a constrained optimization by incorporating the constraints;
$m$ = number of a priori settings associated with climate outlook information to be used to constrain the operational hydrology outlook;
$N$ = number of duplicated scenarios in the hypothetical very large structured set used for statistical estimation in the operational hydrology outlook;
$N_L$ = number of duplicated scenarios, in the hypothetical very large structured set used for statistical estimation in the operational hydrology outlook, which have $T_{JJA} \le \tau_{JJA,0.333}$ in Appendix I;
$N_U$ = number of duplicated scenarios, in the hypothetical very large structured set used for statistical estimation in the operational hydrology outlook, which have $T_{JJA} > \tau_{JJA,0.667}$ in Appendix I;
$n$ = number of scenarios available for use in generating the operational hydrology outlook;
$n_L$ = number of scenarios, available for use in generating the operational hydrology outlook, which have $T_{JJA} \le \tau_{JJA,0.333}$ in Appendix I;
$n_U$ = number of scenarios, available for use in generating the operational hydrology outlook, which have $T_{JJA} > \tau_{JJA,0.667}$ in Appendix I;
$P[\,]$ = probability of the event in brackets;
$\hat{P}[\,]$ = relative frequency in a set, of the event in brackets, used as a probability estimate;
$Q_g$ = total precipitation over period $g$;
$q_{g,i}$ = total precipitation in period $g$ of scenario $i$;
$r_i$ = duplication count for $i$th scenario in the original set of possible future scenarios for the hypothetical very large structured set;
$S^2$ = estimate of variance for variable $X$;
$T_g$ = average air temperature over period $g$;
$t_{g,i}$ = average air temperature in period $g$ of scenario $i$;
$w_i$ = weight applied to $i$th scenario in the original set of possible future scenarios for calculation of statistics for an operational hydrology outlook;
$X$ = a meteorological or hydrological variable;
$x_k^N$ = value for variable $X$ in $k$th duplicated scenario in the hypothetical very large structured set of $N$ scenarios;
$x_i^n$ = value for variable $X$ in $i$th scenario in the original set of $n$ possible future scenarios;
$\bar{x}$ = estimate of mean for variable $X$;
$y_m^N$ = $m$th ordered value for variable $X$, corresponding to $k$th duplicated scenario in the hypothetical very large structured set of $N$ scenarios;
$y_j^n$ = $j$th ordered value for variable $X$, corresponding to $i$th scenario in the original set of $n$ possible future scenarios;
$\theta_{g,\gamma}$ = reference total precipitation $\gamma$-probability quantile for period $g$;
$\lambda_k$ = Lagrange multiplier, representing the penalty associated with violation of the $k$th constraint equation in the optimization;
$\xi_\gamma$ = reference $\gamma$-probability quantile for variable $X$;
$\tau_{g,\gamma}$ = reference average air temperature $\gamma$-probability quantile for period $g$; and
$\Omega$ = set of indices of scenarios.